Symbolic and Motor Contributions to Vocal Imitation in Absolute Pitch
Author(s): Sean Hutchins, Stefanie HutkaSylvain Moreno
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 32, No. 3 (February 2015), pp. 254-265
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/10.1525/mp.2015.32.3.254
Accessed: 17/02/2015 12:15

Symbolic and Motor Contributions to Vocal Imitation in Absolute Pitch

Sean Hutchins
The Royal Conservatory of Music, Toronto, Canada

Stefanie Hutka & Sylvain Moreno
Rotman Research Centre at Baycrest Hospital & University of Toronto, Toronto, Canada

The Linked Dual Representation model (Hutchins & Moreno, 2013) was designed to provide an account for the broad pattern of relationships between vocal perception and production, including both correlations and dissociations between the two. This model makes a unique prediction that musicians with absolute pitch (AP) should be biased towards compensating for objectively mistuned notes in a single note imitation task. In this paper, we tested this prediction by asking musicians with and without AP to imitate vocal notes that are either well-tuned or mistuned. We found that AP musicians were more likely to bias their responses to compensate for mistunings, and that this effect was stronger after longer response delays. We also showed evidence for some implicit AP-like abilities among non-AP musicians. Our findings were predicted by the Linked Dual Representation model, but not other models, providing further evidence for this model.

Received: June 10, 2014, accepted September 27, 2014.

Key words: absolute pitch, singing, tuning, perception and production, Linked Dual Representation model

The act of imitating a vocal pitch, while it may seem simple, yields an important question for researchers studying music production – namely, what is the relationship between perception and production? In order to imitate a tone, we need to perceive that tone, including its low-level features, and then construct a vocal-motor representation of that tone (a representation of how one would move to create that tone with their own voice). In contrast, in order to identify or make a judgment about a perceived tone, we need a symbolic representation of that tone. The symbolic and vocal-motor representations are two different ways of encoding relevant information about a tone, which each support different types of responses (much as two different types of maps may help guide different types of actions). Here, a symbolic representation encodes knowledge *about* a tone (including relevant general semantic knowledge and specific episodic knowledge, such as "This is the tonic tone," or "This is the same as the previous tone"), whereas a vocal-motor representation encodes knowledge of *how to make* a tone (akin to procedural memory). These two representations are influenced by different memory systems, and draw on both long- and short-term memory from these systems when encoding the musical information we perceive.

Figure 1 (left section) presents three potential models describing possible relationships between perception and production abilities. A perception-based model (Figure 1a) posits that low-level perception is first encoded as a symbolic representation, which is then used to create a vocal-motor representation. A motor model (Figure 1b) posits the reverse account; that low-level perception is first encoded as a vocal-motor representation, which is the basis of a symbolic representation. Finally, a dual-route model (Figure 1c) posits that symbolic representations and vocal-motor representations comprise two different pathways, and that neither one is used to create the other. These models are not meant to outline specific brain regions, but rather informational architectures, about the possible flow of information and how it is encoded (see Hutchins & Moreno, 2013, for further information on this topic).

For a long time, it was presumed that this ability to create a symbolic representation of a tone was directly linked to the ability to create a vocal-motor representation of the same tone, as in the perceptual-based and motor models. It is natural to presume that poor singers, for example, have poor perceptual abilities at the root of their problems. Researchers searched for a correlation between vocal perception and production, with varying degrees of success. Some have found such a correlation (e.g., Amir, Amir, & Kishon-Rabin, 2003; Estis, Coblentz, & Moore, 2009; Estis, Dean-Claytor, Moore, & Rowell, 2011; Moore, Keaton, & Watts, 2007; Watts, Moore, & McCaghren, 2005), but others, using similar

Music Perception, volume 32, issue 3, pp. 254–265, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2015.32.3.254

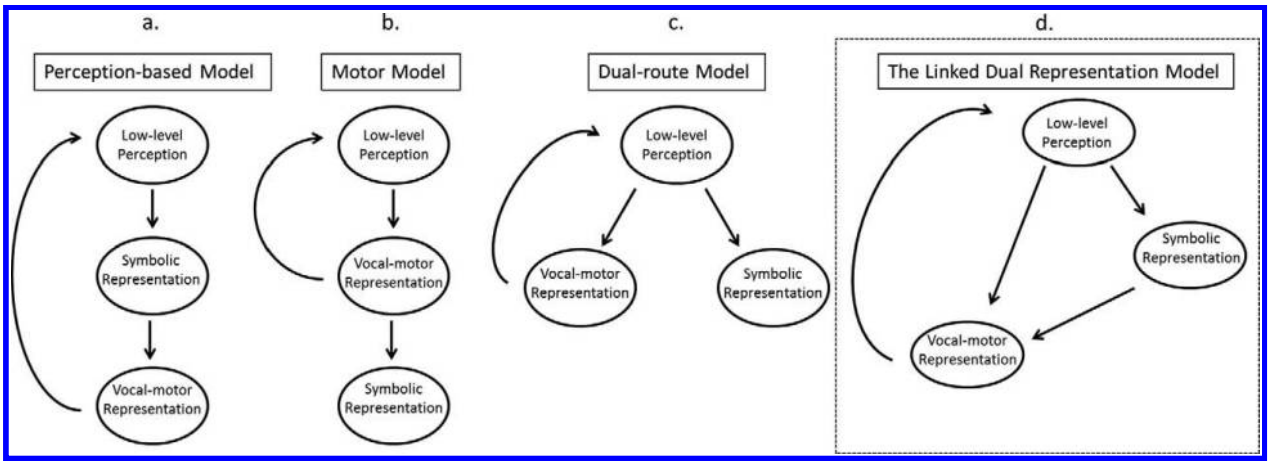

FIGURE 1. Four models of the relationship between perception and production: A perception based model (a), a motor model (b), a dual-route model (c), and the Linked Dual Representation model (d; Hutchins & Moreno, 2013).

designs, have not (e.g., Bradshaw & McHenry, 2005; Dalla Bella, Giguère, & Peretz, 2007; Moore, Estis, Gordon-Hickey, & Watts, 2008; Pfordresher & Brown, 2007).

However, recent research with a special population of participants with congenital amusia has mostly supported the idea that vocal perception and production abilities are, in fact, dissociated. Congenital amusia, though encompassing many different types of musical disabilities, is most commonly manifested as an impairment in conscious pitch perception ability. Because of this impairment, congenital amusia is an interesting test of the relationship between perception and production. Several studies have shown, using a variety of methods, that a subset of congenital amusics (though not all) have a spared production ability; they are able to imitate pitches with their voice to a much higher degree than would be expected if their production abilities were entirely dependent on their overt perceptual abilities (Dalla Bella, Giguère, & Peretz, 2009; Hutchins & Peretz, 2012a, 2013; Loui, Guenther, Mathys, & Schlaug, 2008).

Based on this evidence, as well as some similar types of evidence in unimpaired people (e.g., Hafke, 2008; Vurma, 2010), some researchers have proposed a dual-route model of pitch perception and production (Griffiths, 2008; Hutchins, Zarate, Zatorre, & Peretz, 2010). This type of model suggests that pitch perception and production are each supported by different pathways, allowing for either ability to be impaired without affecting the other. Such a model has been supported by neurological evidence as well, with EEG (Moreau, Jolicoeur, & Peretz, 2009, 2013; Peretz, Brattico, Järvenpää, & Tervaniemi, 2009), fMRI (Hyde, Zatorre, & Peretz, 2011), and DTI (Loui, Alsop, & Schlaug, 2009) studies all suggesting that it may be the higher-, rather than lower-level perceptual pathways that are impaired in amusia, and that production processing could be unaffected. This dual route model also provides an explanation of why perception and production abilities do not always seem to align well in an unimpaired population – why poor singers do not necessarily have worse pitch judgment skills, for example. However, such a model leaves unexplained the weak but persistent correlation found between perception and production across several studies (Amir et al., 2003; Estis et al., 2009; Estis et al., 2011; Hutchins & Peretz, 2012a; Moore et al., 2007; Watts et al., 2005), as well as the production impairments seen in many cases of congenital amusia (Dalla Bella et al., 2009; Hutchins et al., 2010).

Hutchins and Moreno (2013) reviewed the evidence for and against each of these types of models, and concluded that, while each could account for some of the evidence on perception-production relationships, none could provide a framework to account for the broad tendencies in the data – the dissociability of pitch perception and production combined with the weak correlation between the two abilities. Based on this, we proposed a new model, the Linked Dual Representation model (Figure 1d), which could account for both of these tendencies in the data. This model proposes that low-level perceptual information can be processed in two separate ways, such that there are separate pathways for encoding a note as a symbolic representation or as a vocal-motor representation, similar to a dual route model. However, the Linked Dual Representation model also posits that the vocal motor representation can be influenced by the symbolic representation. The identification and judgments of notes can influence the production thereof, under the right circumstances.

The Linked Dual Representation model not only accounts for the current evidence on perception and production of the voice, but also makes some unique predictions. One of these predictions concerns people with absolute pitch (AP). AP is the rare ability, found almost exclusively in musicians, to identify or produce specific musical pitches without reference to another musical pitch (Takeuchi & Hulse, 1993). An AP possessor would be able to provide a label for a pitch they heard, whereas a non-AP possessor would not, even despite extensive music training in many cases. AP does not seem to be due to better basic low-level perceptual abilities – they have similar abilities to non-AP possessors to resolve frequency, spatial location, and temporal differences (Fujisaki & Kashino, 2002). Rather, AP is built upon the higher-level ability to form a stable symbolic representation of a pitch. Although it has proven very difficult to teach AP to adults, there is evidence that childhood experience is a necessary, if not sufficient condition to develop this ability (Levitin & Rogers, 2005; Russo, Windell, & Cuddy, 2003; Takeuchi & Hulse, 1993). Neurologically, this ability is manifested in a hemispheric asymmetry in AP possessors, such that they show larger size and increased use of the left versus right planum temporale than do non AP possessors (Brancucci, di Nuzzo, & Tommasi, 2009; Keenan, Thangaraj, Halpern, & Schlaug, 2001; Zatorre, Perry, Beckett, Westbury, & Evans, 1998). AP possessors also show higher local connectivity in perisylvian brain regions (Jäncke, Langer, & Hänggi, 2012; Loui, Li, Hohmann, & Schlaug, 2011).

AP is not an all-or-none phenomenon; rather, it can be described as a gradient of ability (Bermudez & Zatorre, 2009). On the high end, some with a strong AP ability can be accurate to within 20 cents in their pitch identifications (100 cents = 1 semitone; Miyazaki, 1988; van Krevelen, 1951), and can even automatically identify the pitches of non-instruments, such as car horns, glasses, and spoons (Bachem, 1937). This strong AP ability has been labeled "AP-1" by Baharloo, Johnston, Service, Gitschier, & Freimer (1998). On the other end of the spectrum, some residual AP-like abilities can be found even among non-AP musicians, called "implicit AP" (Miyazaki, 2004; Takeuchi & Hulse, 1993). For example, non-AP possessors tend to show a high degree of consistency when singing songs (Halpern, 1989; Levitin, 1994) or notes (Hsieh & Saberi, 2008), and tend to notice when familiar songs or sounds are played in different keys (Schellenberg & Trehub, 2003; Smith & Schmuckler, 2008; Terhardt & Seewann, 1983). Baharloo et al. (1998) have categorized such individuals with poorer – albeit significantly above-chance – performance on pure-tone AP tests as "AP-2," "AP-3," etc. Individuals with AP-2 are defined by heightened tonal memory (HTM) ability, as contrasted with AP-1 possessors, who have the ability to perceptually encode (APE) the frequency of auditory stimuli (Loui et al., 2011; Ross, Gore, & Marks, 2005). At the neural level, AP-1 possessors have a higher volume of left hemispheric white matter tracts connecting regions central to perception and categorization (i.e., superior temporal gyrus and middle temporal gyrus), as compared to AP-2 and non-AP groups (Loui et al., 2011). In addition, AP-1 possessors have stronger small-world network connectivity as compared to AP-2 and non-AP groups (Loui, Zamm, & Schlaug, 2012).

Even within individual AP possessors, different features of a note can lead it to be more or less easily recognized. AP possessors tend to be more accurate in identifying notes played by their primary instrument (Brammer, 1951; Sergeant, 1969), and most are more accurate at identifying pitches played on the piano than other instruments (Lockhead & Byrd, 1981; Marvin & Brinkman, 2000), possibly due to greater familiarity with that instrument. They also tend to be worse at identifying the pitch of vocal tones than other instruments (Vanzella & Schellenberg, 2010), an effect that may be related to the vocal generosity effect (Hutchins, Roquet, & Peretz, 2012), in which all listeners tend to be worse at determining the tuning of vocal pitches. There is also a general effect for AP possessors to be better at identifying white-key, rather than black-key pitches (which tend to be less frequent; Dohn, Garza-Villarreal, Ribe, Wallentin, & Vuust, 2014; Miyazaki, 1988), again highlighting the role of experience in AP ability. The relevance of any and all of these categories varies between AP possessors, and AP itself can vary from being highly automatic to barely present.

The ability of AP possessors to label individual tones can be thought of as a type of symbolic representation ability, as it constitutes knowledge about a note (see Figure 1). As such, it provides an interesting way to test the predictions of the Linked Dual Representation model (Hutchins & Moreno, 2013) against the other models. This model predicts that a singer imitating a single note should primarily make use of the pathways directly from low-level perception to the vocal-motor representation, especially for faster responses. However, forcing a delayed response should lead to greater mediation from the singers' symbolic representation of the note. In most cases, this would not make much of a difference on the final pitch of the produced note, especially for individual notes. Even in cases where the stimulus note is objectively out-of-tune (sharp or flat relative to standard equal temperament tuning), most singers would not be able to determine this, and should be uninfluenced by this objective tuning. However, in the case of AP possessors, their ability to categorize and label individual notes would make them able to provide such a label. In cases of delayed responses, then, they should be influenced by the name of the note (which is a symbolic representation) as well as the pitch they heard. Thus, their vocal response should be biased to be closer in line with perfectly-tuned notes, especially after a delay. For example, if an AP possessor hears a tone that is a D, 20 cents flat, their vocal-motor response should be directly influenced by this precise pitch. However, if they label it as "a type of D," then their vocal-motor response should also be influenced by this label, and their imitative response should be sharper than it otherwise would have been.

This response bias among AP possessors is not predicted by the other models. A dual route model predicts that any symbolic categorization should have no influence on a vocal-motor representation. A motor model makes the same prediction. A perception-based model predicts that the symbolic representation should wholly determine the vocal-motor representation, and thus the response bias should not be a function of the response delay.

In this study, we tested the unique predictions of the Linked Dual Representation model by asking highly trained musicians, either with or without AP, to listen to individual tones and then imitate them as closely as possible. These tones were either perfectly in-tune or mistuned sharp or flat. Participants sang back the tones either immediately after hearing them, or after delays from 2-10 seconds. In addition, these delays could either be silent or filled with white noise. If the Linked Dual Representation model is correct, then AP participants should show a greater tendency than non-AP participants to bias their imitative responses in the direction of the well-tuned note (sharp responses for objectively flat tones, and flat responses for objectively sharp tones), to compensate for the mistuning. Furthermore, this compensatory response bias should be mediated by the enforced response delay, with longer delays leading towards stronger biases. Filling this response delay with white noise may also have the effect of heightening this response bias. We also expect lower overall accuracy for all participants in longer, filled response delays (Estis et al., 2009; Estis et al., 2011).

Method

PARTICIPANTS

The participants were 37 highly trained musicians (20 male), recruited mainly from The Royal Conservatory of Music Glenn Gould School and the University of Toronto music program. These were recruited from advertisements calling for participants "with or without perfect pitch." Participants had an average age of 22.97 years ($SD = 3.63$), and had been training on their *primary* instrument for an average of 13.11 years ($SD = 4.36$; with most being proficient on other instruments as well). There was a range of instruments played, with violin and voice being the two most represented among the primary instruments. Among those later categorized as AP possessors (13 in total, 6 male, see below), there was a trend toward more years of training on their primary instrument (15.66 years for AP, 11.83 years for non-AP), though this was not significant, $t(25) = 1.92$, $p = .07$. This was likely due to the greater propensity for piano and violin among AP possessors, as these are the instruments in which children are likely to begin playing (people who play instruments such as the tuba or saxophone tend not to begin playing on these instruments until later in age; many of these participants had begun their training earlier on a different instrument). Most of those whose primary instrument was the voice were non-AP possessors (seven of eight); all of these participants indicated at least moderate proficiency on other instruments as well. All participants indicated at least moderate levels of vocal proficiency.

STIMULI AND PROCEDURE

As stimuli for the AP test, we created a set of 20 pure tones, sampled evenly from the range of equal-tempered tones (corresponding to a piano key) between A2 (110 Hz) and B5 (988 Hz), so chosen because they encompass a range within that of commonly sung vocal pitches. Pure tones were chosen so as not to bias towards any instrument, and make it a fairly difficult test of AP. Every pitch class was represented at least one time, and no individual pitch was repeated. We also created a set of 40 additional pure tones that were slightly sharp or flat with respect to the original 20 pure tones. Of this set of 40, ten tones were 40 cents sharp of perfect tuning, ten were 20 cents sharp, ten were 20 cents flat, and ten were 40 cents flat. Each tone lasted one second, with a 100 ms ramped onset and offset. The original 20 tones were arranged into a single pseudo-randomized list, such that pitch classes could not repeat within a span of three items. These 20 tones comprised the stimuli for the first section of the AP test. Another pseudo-randomized list was created including all 60 tones, with the same constraints. These 60 tones comprised the stimuli for the second section of the AP test.

To create the stimuli for the main singing experiment, we recruited a trained male and female singer to produce tones at five different pitch levels: C, D, E, F#, and G# (C3-G#3 for the male, C4-G#4 for the female), all on the syllable /du/ ("doo"). These were chosen so as to not evince a particular tonality, and to lie within the vocal range of most people. Each was performed with minimal vibrato, for approximately 1.5 s. Each of these recordings was then imported into Melodyne (Celemony Software GmbH, Munich Germany), where it was first manipulated to make the duration of the clip equal 1.5 s total, including 100 ms of silence at the beginning and end of each clip. Then, we used Melodyne to correct minor tuning errors in the clips, by correcting the mean pitch to equal tempered tuning (A = 440 Hz) and removing variability in the pitch of the vocal signal. Thus, each of the 10 clips were pure examples of a single pitch with a vocal timbre. Following this, we created pitch-shifted version of each of these clips, at pitches -40, -20, +20, and +40 cents from the original modified clip – each shifted version retains the timbre and amplitude variations of the original, while changing only in pitch. This created 25 male-voice stimuli and 25 female-voice stimuli.

We also used Adobe Audition (Adobe Systems, San Jose, US) to create white noise or silence lasting .01, 2, 5, and 10 s. The white noise was given a 100 ms ramped onset and offset and normalized for amplitude to -3 dB.

PROCEDURE AND DESIGN

Participants first participated in a test of AP ability. In this test, they heard one of the pure tones, and were then shown a screen with 48 chromatic notes on a set of four staves, ranging from C2 to B5. Participants were asked to click on the note that matched the one they heard, with no opportunities to listen a second time. All participants were presented with the same list of 20 tones, including only those with an equal-tempered pitch, presented in a pseudo-randomized order, such that pitch classes could not repeat within a span of three items. If participants correctly identified 25% of the tones, they were asked to continue to the next section of the AP test; otherwise, they immediately began the singing experiment. Three practice trials including only text (no sounds; e.g., "Click on C4 [Middle C]") were included in the beginning of the experiment, in order to familiarize the participants with the task.

In the second section of the AP test, the task was largely similar. However, in this section, there were 60 trials all in all, and these included both perfectly tuned tones as well as the mistuned versions in pseudo-random order (as above). Participants were informed that some examples would be mistuned, and were asked to click on the note that was closest to the tone they heard. Following this, a display box with five options popped up, and participants were asked to decide whether they thought the note that they heard was very flat, somewhat flat, in tune, somewhat sharp, or very sharp.

After the AP test was completed, participants began the singing imitation task. Here, participants heard one of the 25 vocal stimuli of their own gender, and were instructed to imitate this tone as closely as possible. They were informed that only pitch would vary. Immediately following each vocal stimulus, we played one of the filler stimuli (white noise or silence, lasting .01, 2, 5, or 10 s). This varying response delay was followed by the text "Respond Now" appearing on the screen. Participants were instructed not to respond until this text appeared on the screen, and were instructed not to subvocalize the tone in the interim. Each stimulus was presented with each level of Response Delay (4 levels), and Filler (Noise vs. Silence) one time, leading to 200 trials in the experiment. These were presented in one of four pseudo-randomized lists, in which base notes could not be immediately repeated (thus, all tones were a minimum of 120 cents from the previous tone).

Before coming in, participants completed a 10-min questionnaire on their music and language background. The AP task lasted approximately 5 min for those without AP, and approximately 10-15 min for those with AP. The singing imitation task lasted about 35 min in total. All tasks were controlled using ePrime (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania, USA). All stimuli were heard through open BeyerDynamic DT 770 Pro headphones (Beyerdynamic GmbH & Co. KG, Heilbronn, Germany), and imitations were recorded with a Sennheiser MKE 2 microphone (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany), through Reaper software (Cockos, San Francisco, USA).

ANALYSIS

To measure AP presence and strength, we calculated the mean octave-corrected absolute error (hereafter, AP error) for each participant. This is a measurement of how far off each participants' average pitch identification score in the AP test was, in semitones – thus low scores indicate stronger AP ability. First, we corrected each identification response to the nearest octave register. We then measured the absolute value of the difference between this and the correct response. For judgments in the second part of the AP test, we used the base response, and converted the "very flat, somewhat flat, in tune, somewhat sharp, or very sharp" responses into -40 cents, -20 cents, 0 cents, +20 cents, and +40 cents modifiers, respectively. These were then compared to the pitch of the correct response. This allows for precisely correct responses to these tones, and possible identification errors in steps of 20 cents. In this metric, errors of more than 6 semitones are not possible (due to octave correction), and thus the mean AP error would tend towards 3 semitones, for totally random identification judgments. AP error was calculated from an average of both sections of the experiment, as it was found that participants with any measurable AP ability performed similarly across both sections of the AP test (and non-AP participants tended toward chance in the first section).

Responses in the singing task were analyzed for pitch in Melodyne. We calculated the mean pitch of the response, and used the difference between this and the target tone as the response error (a signed measurement). We also calculated the absolute value of this measurement, as a measurement of generalized singing error (as in Hutchins & Peretz, 2012b). Errors were omitted from further analysis – this included responses before the appropriate time, doubled responses, no responses, and unanalyzable signals. These comprised 0.38% of all trials.

Results

AP TEST

We grouped participants into AP and Non-AP (NAP) categories based on a threshold of less than 3 standard deviations from the mean of self-identified NAP participants. Most trained musicians are aware of whether they have AP or not, though we did identify two clear cases of AP where the musicians were previously unaware of this ability; no NAP musicians (as determined by the AP test) had indicated that they unambiguously had AP. This cut-off yielded a threshold of 1.47 semitones AP error. However, it should be noted that there was a clear break between AP and NAP participants, as all AP participants showed AP error scores of less than 1 semitone, whereas the NAP participants were no lower than 1.7 semitones AP error. These results showed that 13 of our participants were clear AP possessors, whereas the remaining 24 were categorized as NAP. Analyzing using only base note data (omitting fine-grained tuning information) did not change the shape of these data. No AP participants showed evidence of a systematic tuning bias greater than half a semitone, indicating that their AP labelling was aligned with normal tuning.

We found no overall correlation between response time and AP error, $r(35) = .13$, *ns* However, when analyzing solely within participants shown to have AP, the correlation between reaction time and AP error was highly significant, $r(11) = .71$, $p < .01$. Participants with lower overall errors in the AP test were also faster to respond, likely indicating a more automatic AP ability.

SINGING IMITATION TASK

To measure whether AP participants are more likely to compensate for tuning deviations in their vocal responses (responding more towards perfect tuning), we ran a 2 x (5 x 4 x 2) mixed-design ANOVA over the factors of AP, (AP vs. NAP), Tuning Deviation (-40, -20, 0, +20, +40), Response Delay (immediate [.01 s], 2 s, 5 s, and 10 s), and Filler (noise vs. silence), using response error as our dependent variable. We found significant main effects of Response Delay, $F(3, 105) = 18.63$, $p < .001$, η^2 generalized = .02, and Tuning Deviation, $F(4, 140) = 13.88$, $p < .001$, η^2 generalized = .04 and a significant interaction between Response Delay and Tuning Deviation, $F(12, 420) = 5.92$, $p < .001$, η^2 generalized = .01. In addition, we also found a significant interaction between AP and Tuning Deviation, $F(4, 140) = 2.91$, $p = .02$, η^2 generalized = .009. Figure 1 shows a graph of response error for AP and NAP participants, across each Tuning Deviation, for all four Response Delays. As was predicted, AP participants tend to show evidence of compensation for mistunings in their vocal responses, with sharper responses to objectively flat tones (-40 and -20), and flatter responses to objectively sharp tones (+20 and +40). Their tendency to compensate for these errors was greater after larger response delays. However, interestingly, we found evidence for this tendency among NAP participants as well. NAP participants showed the same tendency for compensation of mistunings as AP participants, though not to the same extent. Because of this, the expected three-way interaction between AP, Tuning Deviation, and Response Delay did not reach significance, even though the results trended in this direction. The factor of Filler did not have a main effect or any significant interactions. No other main effects or interactions reached significance. Participants were very accurate overall, with mean absolute errors near 16 cents. There was a very small effect (2 cents) of increasing time delay, and no effects of noise or any other variables.

To follow up on the interaction between AP and tuning deviation, we examined the role of the strength of the participant's AP on the degree to which they showed

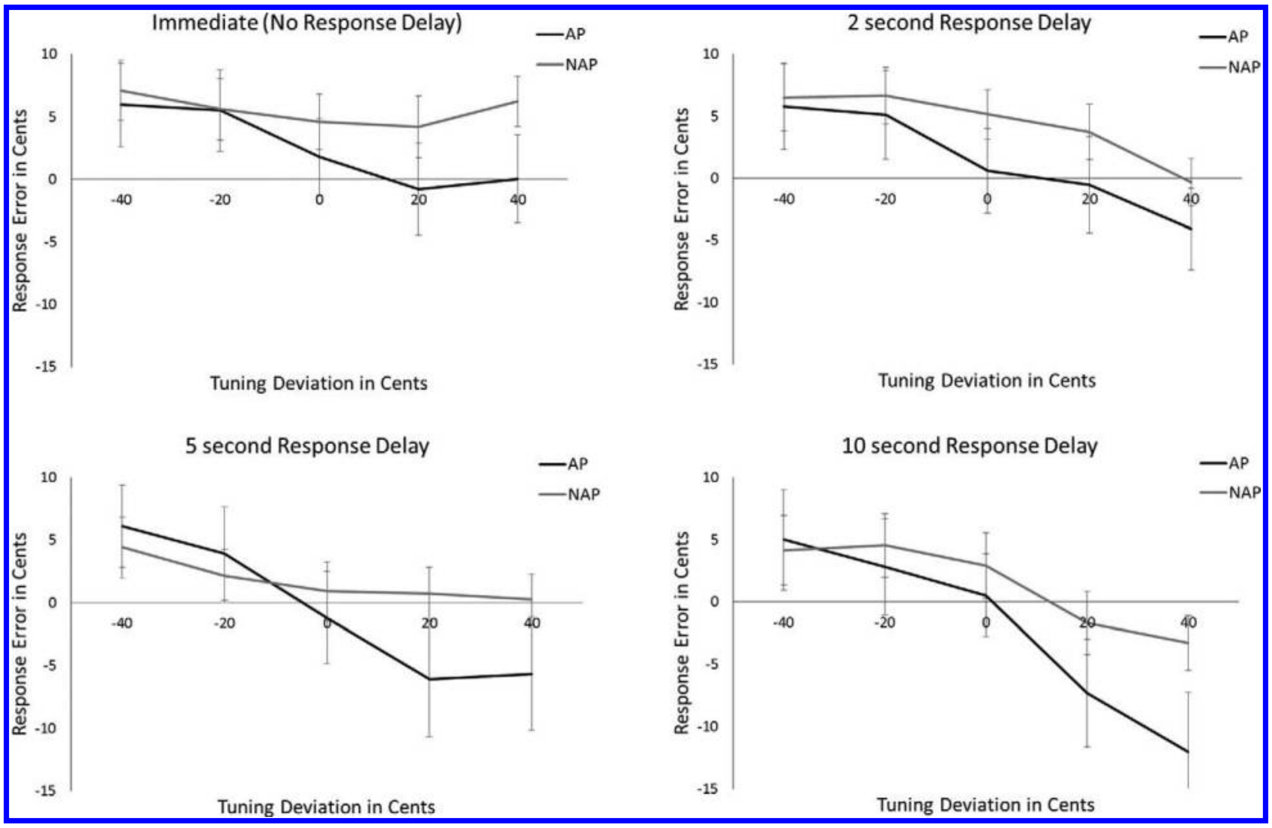

FIGURE 2. The average response errors of AP and NAP participants for each tuning deviation, across the four different response delays, with standard error bars.

compensatory tuning in their responses. To do this, we computed the degree to which each individual response to mistuned notes was compensated back towards the perfectly tuned note, in terms of percentage (responses to perfectly tuned notes were excluded from this analysis). Here, a response 40 cents higher than a note tuned 40 cents flat indicates 100% compensation, and a response 10 cents lower than a note tuned 20 cents sharp indicates 50% compensation. This score can be greater than 100%, or negative as well, for overcompensations or additive errors. We correlated each participant's average percent compensation with their AP error from their AP test. Participants with lower AP error scores tended to show greater compensation for mistuning in their responses, $r(35) = -.36, p < .05$. Figure 2 shows this relationship.

In examining this correlation, it became clear that among AP participants, there were two distinct groups. One group was a strong AP group (i.e., AP-1; 7 participants), with AP error scores lower than 0.33 (less than 33 cents of error in an average judgment), whereas the other group was a weak AP group (i.e., AP-2, or HTM; 6 participants), with error scores between 0.67 and 1.0 (average error of less than a semitone). There were no cases of AP errors between 0.33 and 0.67. It also became clear that the strong AP group seemed to show a stronger compensatory response than the weak AP group. The strength of the correlation rose slightly when analyzed within only AP participants. Based on this, we performed a reanalysis of our data, in which we split our AP groups into strong and weak AP categories (making three total groups). When divided in this way, no members of any group were within 3 standard deviations of another group. Similar divisions of participant groups into strong AP (i.e., AP-1; APE), weak AP (i.e., AP-2; HTM), and non-AP categories have previously been employed in the AP literature (e.g., Miyazaki, 1990; Ross et al., 2005; Loui, Li, et al. 2011; Loui, Zamm, & Schlaug, 2012).

We tested a 3 x 4 mixed design ANOVA with the factors of AP category and Response Delay, using percent compensation as the dependent variable. This analysis showed a significant effect of AP category, $F(2, 34) = 3.62$, $p = .04$, η^2 generalized $= .15$, and a significant

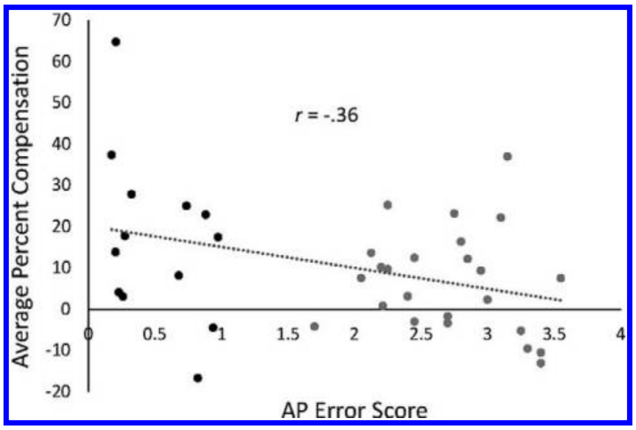

FIGURE 3. The relationship between average AP error score and average percent compensation, for each participant, with the regression line. AP participants are marked in black, NAP participants are marked in grey.

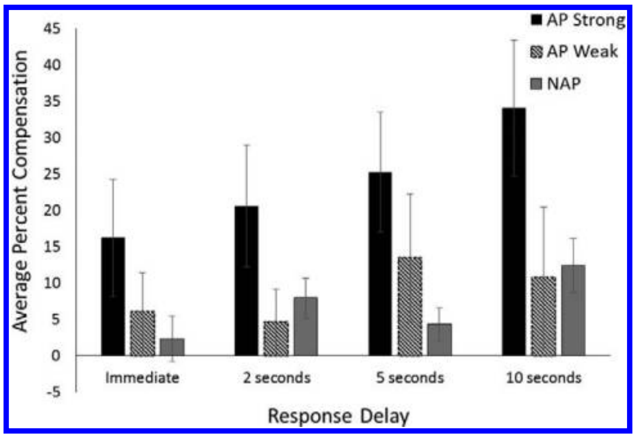

FIGURE 4. The average percent compensation of strong AP, weak AP, and NAP participants at each response delay, with standard error bars.

effect of Response Delay, $F(3, 102) = 9.64$, $p < .001$, η^2 generalized = .05. These factors are shown in Figure 3. As can be seen, the strong AP group showed the largest propensity to compensate for tuning errors in their response, and this tendency became more pronounced after longer delays. The weak AP group showed a pattern of response more similar to the NAP group. However, the trend for an interaction between AP category and Response Delay did not reach significance, likely due to the relatively small number of strong AP participants.

In order to determine whether patterns in relative, rather than absolute pitch, could be driving the weakly compensatory pattern in NAP participants, we computed the intervallic distance from each stimulus tone to the next, and computed the deviations between each interval and the nearest perfectly tuned interval. We correlated these intervallic deviations with the absolute deviations, and found no hint of a correlation, $r(7361) = .026$, *ns*. This indicates that sharp or flat intervals were equally likely to occur on sharp or flat notes; thus intervallic deviations cannot explain the pattern in our data.

Discussion

Our data showed that, when presented with a tone that was objectively flat in comparison to standard tuning, AP participants' imitative responses were more likely to be sharp, and when presented with tones that were sharp, their responses were more likely to be flat. This compensatory response bias was stronger across longer delays. This pattern of effects is predicted by the Linked Dual Representation model (Hutchins & Moreno, 2013), which posits that AP possessors will show a tendency for their AP ability to influence singing imitation responses, such that they compensate for errors in tuning. This effect is predicted to be especially pronounced when the response was delayed.

This response pattern supports the idea that, among people with AP, the imitative response (determined by the vocal-motor representation) was not only directly influenced by the low-level perceptual encoding, but also by the symbolic representation of the note. Musicians with AP have the ability to identify individual tones without a reference pitch, and thus can symbolically represent the name of a note. When musicians with AP hear a note to imitate, then, they can be influenced by the name of the note as well as the pitch they heard. For example, an E, 20 cents sharp, would be encoded not just as "the note I make in this way with my voice," but also as "An instantiation of 'E.'" This would tend to bring their responses closer in line with perfectly tuned notes.

Among the possible models of vocal-motor and symbolic encoding, the Linked Dual Representation model can best account for this pattern of effects. A straight dual-route model would predict that the symbolic encoding of a tone would have no effect on a vocal-motor encoding, and thus no compensatory effect on the imitative response. This same prediction would also be made by a motor model. A perceptual-based model would predict this compensatory effect; however, it does not predict the change in this effect across response delays – because all responses are mediated through the symbolic encoding, there is no way for it to become stronger across time. In contrast, in the Linked Dual Representation model, a vocal-motor representation is influenced by both low-level perceptual encoding and a symbolic representation. Since the symbolic representation pathway involves more steps, it can be presumed to take longer to influence the vocal-motor-representation, which predicts more influence from this pathway after longer response delays, as we found in our data.[1]

One unexpected outcome of this experiment was that even musicians without AP showed a weaker version of this same compensatory effect. Because NAP musicians do not have access to labels of notes, we did not expect them to be able to determine whether any individual instance of a note is sharp or flat, relative to perfect tuning. These participants did not claim to have AP, and made no indication that they were aware of their performance biases. However, their responses did show systematic biases that indicated a subconscious awareness of perfect tuning. Although these musicians are highly trained, and have strong relative pitch abilities, our analyses show no unforeseen correlations between mistunings of intervals and individual tones, which can rule out relative pitch as an explanation for this finding.

We can propose two possible interpretations for this finding. The first is an effect of long-term motor memory: Participants should have much more practice singing well-tuned notes than mistuned notes, and thus those motor commands should be more ingrained (see Hsieh & Saberi, 2008, for a similar explanation). Here, the implicit knowledge of pitch categories resides in the memory of vocal-motor representations, rather than in the symbolic representations. However, this explanation seems to predict much less variability in actual singing tuning than we see in our data. It also seems to predict that those musicians who trained as singers should show a stronger effect than those who did not (as they have had considerably more practice with those motor commands). Upon inspection, this did not seem to be borne out in our data, though a true test would require more trained singers than we had available in this data set (there were eight in total) and a slightly different design.

A second explanation is that these highly trained musicians may have had some form of implicit AP ability, an unconscious knowledge of pitch identities that can arise in certain settings or tasks (Levitin & Rogers, 2005; Miyazaki, 2004). This explanation posits an AP ability as part of the symbolic representation, but not highly activated enough to reach the level of explicit knowledge. As part of their training, these participants listen to and perform music for hours every day. It is possible that this near-constant exposure to music, the vast majority of which is well-tuned, could lead to an implicit AP ability that may have affected their imitative responses. Miyazaki (2004) suggests just such an ability in long-term memory for repeatedly heard tones; both Halpern (1989) and Levitin (1994) describe a similar phenomenon among nonmusicians in their singing (though see Frieler et al., 2013), and other studies (Schellenberg & Trehub, 2003; Smith & Schmuckler, 2008; Terhardt and Seewann, 1983) showed evidence for an implicit AP ability in perception. Deutsch (1987) also describes an implicit AP-like response bias to the tritone paradox. This implicit AP ability may be the cause of NAP musicians' compensatory responses, using the same mechanisms as the AP participants, but with a lower activation of the symbolic representation and therefore a weaker compensatory response bias. In this interpretation (but not the motor memory interpretation), the NAP participants' compensatory response biases should be affected by the response delay in the same manner as the AP participants'. We see this effect in the data, which suggests this may be a better interpretation.

Despite this unexpected finding, our data nevertheless show that AP participants showed the compensatory response to a larger extent than the NAP participants. AP and NAP participants did not differ in overall accuracy, rather the overall difference was in the direction of their errors. Follow-up analyses indicated that the extent of the compensatory response was correlated with the strength of the participants' AP ability. The participants who were most accurate with their AP judgments showed the strongest effect on their imitations of mistuned responses. These are the participants who responded most accurately, who were also the ones with the fastest overall responses. It is exactly this group, for whom AP is the most automatic, who should show the biggest effect. We also found a trend for this group to be affected the most by the response delays, though this analysis did not reach significance. Taken together, this provides evidence that the compensatory responses are in fact due to AP ability.

[1] As pointed out by a reviewer, the perceptual-based model (1a) can be adapted to also make this prediction of a change in the compensatory effect across response delays. By assuming that the symbolic representation itself can evolve over time (in this case, becoming more categorical, and less directly related to the original tone), the new dynamic 1a model makes essentially the same predictions with regards to the current data. Such a model also predicts that, at no delay, the vocal-motor encoding would be basically uninfluenced by the symbolic representation, as that has had no time to develop. Essentially, this model uses a dynamic representation in place of two pathways to account for the influence of time. While there is other evidence that can dissociate these models (such as the spared production abilities of a subset of amusics), the present experiment can only differentiate between the static models presented in Figure 1.

Interestingly, although delaying the imitative response made a large difference in the compensatory responses, as well as a small difference to overall error, it did not seem to matter whether this delay was silent or filled with white noise. Although other studies have found that time delays (Estis et al., 2009) and noise (Estis et al., 2011) can increase singing error, our findings seem to run counter to this. We suspect that, for our highly trained participants, white noise did not provide enough of an interference to affect their responses. This type of filler may have been too easy to ignore, especially as it required no overt attention. It is possible that more tonal filler material may have then increased both the compensatory responses and the overall error. Longer delays, as well, may contribute to this pattern, as we saw no signs that participants were reaching some kind of threshold after 10 s.

Our choice to use sung vocal tones as our stimuli to be imitated likely had a modest effect on the overall shape of our results. Prior work shows that people are more accurate at imitating voices than non-voices (Goetze, Cooper, & Brown, 1990; Hutchins & Peretz, 2012b; Lévêque, Giovanni, & Schön, 2012; Watts & Hall, 2008). However, it has also been shown that AP possessors are less accurate at judging vocal tones than non-vocal tones (Vanzella & Schellenberg, 2010). This means that our AP participants would have had more activation from low-level perception directly to their vocal-motor representations, and less activation from their symbolic representations to their vocal-motor representations, making this a fairly conservative test of this compensatory response. We would suggest that using instrumental tones (especially those from the piano or the participants' main instrument) would lead to more automatic AP identification and less accurate vocal responses, increasing the effect size.

The AP test results also help to reveal more about the character of AP itself, and the ways in which AP possessors vary. While a naïve theory might suppose a speed-accuracy tradeoff within AP, we found that those who were more accurate in the AP test were also faster, confirming that AP can vary in its automaticity (Bermudez & Zatorre, 2009; Miyazaki, 1988). Many researchers have speculated that AP exists along a continuum, and even non-AP possessors may have latent AP abilities (Hsieh & Saberi, 2008; Levitin & Rogers, 2005; Miyazaki, 2004). Our results suggest that vocal production may have the potential to reveal implicit AP abilities where more standard pitch identification tests do not. This same pattern, in which vocal production tasks reveal implicit abilities, has been shown in several other fields, such as amusia research (Dalla Bella et al., 2009; Hutchins & Peretz, 2012a, 2013), repetition priming (Hutchins & Palmer, 2008), and pitch shifting (Hafke, 2008; Hutchins & Peretz, 2013). We would suggest that a more targeted test of self-identified NAP participants using vocal production would be able to show a stronger effect of implicit AP.

Conclusions

This experiment was designed to test the predictions of the Linked Dual Representation model (Hutchins & Moreno, 2013). Our results show that the vocal productions of AP possessors are affected by their ability to identify and categorize individual tones, such that they show a compensatory response bias – an effect not predicted by any other model of the relationship between perception and production. We also found evidence for a weaker version of this same effect among non-AP possessors, which may constitute evidence for a low-level implicit AP ability. Overall, this study provides an important piece of evidence for understanding the functioning of the vocal-motor system, as well as the overall relationship between perception and production abilities. We would suggest that comparing imitation and judgment tasks in a normal population may also reveal important information about this perception-production relationship.

Author Note

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to S.M. and an Ontario Graduate Scholarship to S. Hutka. We wish to thank Steven Demorest and Peter Pfordresher for organizing this special edition of *Music Perception*, as well as Simone Dalla Bella, Psyche Loui, Graham Welch, and Joanne Rutkowski for their comments and discussions which helped to clarify the role of the model. We would also like to thank Courtney Smith for her contributions to the testing and comments on the manuscript.

Correspondence concerning this article should be addressed to Sean Hutchins, Director of Research, The Royal Conservatory of Music, 273 Bloor St. W, Toronto, Ontario, Canada, M5S 1V6. E-mail: Sean.Hutchins@rcmusic.ca

References

AMIR, O., AMIR, N., & KISHON-RABIN, L. (2003). The effect of superior auditory skills on vocal accuracy. *Journal of the Acoustical Society of America, 113*, 1102-1108.

BACHEM, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America, 9*, 146-151.

BAHARLOO, S., JOHNSTON, P. A., SERVICE, S. K., GITSCHIER, J., & FREIMER, N. B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics, 62*, 224-231.

BERMUDEZ, P., & ZATORRE, R. J. (2009). A distribution of absolute pitch ability as revealed by computerized testing. *Music Perception, 27*, 89-101.

BRADSHAW, E., & MCHENRY, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice, 19*, 431-439.

BRAMMER, L. M. (1951). Sensory cues in pitch judgment. *Journal of Experimental Psychology, 41*, 336-340.

BRANCUCCI, A., DI NUZZO, M., & TOMMASI, L. (2009). Opposite hemispheric asymmetries for pitch identification in absolute pitch and non-absolute pitch musicians. *Neuropsychologia, 47*, 2937-2941.

DALLA BELLA, S., GIGUÈRE, J. F., & PERETZ, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121*, 1182-1189.

DALLA BELLA, S., GIGUÈRE, J. F., & PERETZ, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America, 126*, 414-424.

DEUTSCH, D. (1987). The tritone paradox: Effects of spectral variables. *Perception and Psychophysics, 41*, 563-575.

DOHN, A., GARZA-VILLARREAL, E.A., RIBE, L. R., WALLENTIN, M., & VUUST, P. (2014). Musical activity tunes up absolute pitch ability. *Music Perception, 31*, 359-371.

ESTIS, J. M., COBLENTZ, J. K., & MOORE, R. E. (2009). Effects of increasing time delays on pitch-matching accuracy in trained singers and untrained individuals. *Journal of Voice, 23*, 439-445.

ESTIS, J. M., DEAN-CLAYTOR, A., MOORE, R. E., & ROWELL, T. L. (2011). Pitch-matching accuracy in trained singers and untrained individuals: The impact of musical interference and noise. *Journal of Voice, 25*, 173-180.

FRIELER, K., FISCHINGER, T., SCHLEMMER, K., LOTHWESEN, K., JAKUBOWSKI, K., & MÜLLENSIEFEN, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae, 17*, 334-349.

FUJISAKI, W., & KASHINO, M. (2002). The basic hearing abilities of absolute pitch possessors. *Acoustical Science and Technology, 23*, 77-83.

GOETZE, M., COOPER, N., & BROWN, C. J. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education, 104,* 16-37.

GRIFFITHS, T. D. (2008). Sensory systems: Auditory action streams? *Current Biology, 18*, R387-R388.

HAFKE, H. Z. (2008). Nonconscious control of fundamental voice frequency. *Journal of the Acoustical Society of America, 123*, 273-278.

HALPERN, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory and Cognition, 17*, 572-581.

HSIEH, I. H., & SABERI, K. (2008). Dissociation of procedural and semantic memory in absolute-pitch processing. *Hearing Research, 240*, 73-79.

HUTCHINS, S., & MORENO, S. (2013). The Linked Dual Representation model of vocal perception and production. *Frontiers in Psychology*, 4, 825. doi: 10.3389/fpsyg.2013.00825

HUTCHINS, S., & PALMER, C. (2008). Repetition priming in music. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 693-707.

HUTCHINS, S., & PERETZ, I. (2012a). Amusics can imitate what they cannot discriminate. *Brain and Language, 123*, 234-239.

HUTCHINS, S., & PERETZ, I. (2012b). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology, 141*, 76-97.

HUTCHINS, S., & PERETZ, I. (2013). Vocal pitch shift in congenital amusia (pitch deafness). *Brain and Language, 125*, 106-117.

HUTCHINS, S., ROQUET, C., & PERETZ, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, *30*, 147-159.

HUTCHINS, S., ZARATE, J. M., ZATORRE, R. J., & PERETZ, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America, 127*, 504-512.

HYDE, K. L., ZATORRE, R. J., & PERETZ, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex, 21*, 292-299.

JÄNCKE, L., LANGER, N., & HÄNGGI, J. (2012). Diminished whole-brain but enhanced peri-sylvian connectivity in absolute pitch musicians. *Journal of Cognitive Neuroscience, 24*, 1447-1461.

KEENAN, J. P., THANGARAJ, V., HALPERN, A. R., & SCHLAUG, G. (2001). Absolute pitch and planum temporale. *Neuroimage, 14*, 1402-1408.

LÉVÊQUE, Y., GIOVANNI, A., & SCHÖN, D. (2012). Pitch-matching in poor singers: Human model advantage. *Journal of Voice*, *26*, 293-298.

LEVITIN, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics, 56*, 414-423.

LEVITIN, D. J., & ROGERS, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences, 9*, 26-33.

Lockhead, G. R., & Byrd, R. (1981). Practically perfect pitch. *Journal of the Acoustical Society of America, 70*, 387-389.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience, 29*, 10215-10220.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology, 18*, R331-R332.

Loui, P., Li, H. C., Hohmann, A., & Schlaug, G. (2011). Enhanced cortical connectivity in absolute pitch musicians: A model for local hyperconnectivity. *Journal of Cognitive Neuroscience, 23*, 1015-1026.

Loui, P., Zamm, A., & Schlaug, G. (2012). Enhanced functional networks in absolute pitch. *Neuroimage, 63*, 632-640.

Marvin, E. W., & Brinkman, A. R. (2000). The effect of key color and timbre on absolute pitch recognition in musical contexts. *Music Perception*, *18*, 111-137.

Miyazaki, K. I. (1988). Musical pitch identification by absolute pitch possessors. *Perception and Psychophysics, 44*, 501-512.

Miyazaki, K. I. (1990). The speed of musical pitch identification by absolute-pitch possessors. *Music Perception, 8*, 177-188.

Miyazaki, K. I. (2004). How well do we understand absolute pitch? *Acoustical Science and Technology, 25*, 426-432.

Moore, R. E., Estis, J., Gordon-Hickey, S., & Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice, 22*, 399-407.

Moore, R. E., Keaton, C., & Watts, C. (2007). The role of pitch memory in pitch discrimination and pitch matching. *Journal of Voice, 21*, 560-567.

Moreau, P., Jolicœur, P., & Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Annals of the New York Academy of Sciences, 1169*, 191-194.

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition, 81*, 337-344.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: In tune, out of key, and unaware. *Brain, 132*, 1277-1286.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception, 25*, 95-115.

Ross, D. A., Gore, J. C., & Marks, L. E. (2005). Absolute pitch: Music and beyond. *Epilepsy and Behavior, 7*, 578-601.

Russo, F. A., Windell, D. L., & Cuddy, L. L. (2003). Learning the "special note": Evidence for a critical period for absolute pitch acquisition. *Music Perception, 21*, 119-127.

Schellenberg, E. G. & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14*, 262-266.

Sergeant, D. (1969). Experimental investigation of absolute pitch. *Journal of Research in Music Education*, *17*, 135-143.

Smith, N. A., & Schmuckler, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch processors. *Journal of the Acoustical Society of America, 123*, EL77-EL84.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*, 345-361.

Terhardt, E., & Seewann, M. (1983). Aural key identification and its relationship to absolute pitch. *Music Perception*, *1*, 63-83.

van Krevelen, A. (1951). The ability to make absolute judgments of pitch. *Journal of Experimental Psychology, 42*, 207-215.

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: Effects of timbre on note-naming ability. *PloS One*, 5, e15449.

Vurma, A. (2010). Mistuning in two-part singing. *Logopedics Phoniatrics Vocology, 35*, 24-33.

Watts, C. R., & Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. *Logopedics Phonatrics Vocology, 33*, 74-82.

Watts, C. R., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice, 19*, 534-543.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences*, *95*, 3172-3177.